基于核学习方法的短时交通流量预测 *

王秋莉，李　军

(兰州交通大学 自动化与电气工程学院，兰州 730070)

摘　要：基于核学习的强大非线性映射性能，针对短时交通流量预测，提出一类基于核学习方法的预测模型。核递推最小二乘(KRLS)基于近似线性依赖(approximate linear dependence，ALD) 技术可降低计算复杂度及存储量，是一种在线核学习方法，适用于较大规模数据集的学习；核偏最小二乘(KPLS)方法将输入变量投影在潜在变量上，利用输入与输出变量之间的协方差信息提取潜在特征；核极限学习机(KELM)方法用核函数表示未知的隐含层非线性特征映射，通过正则化最小二乘算法计算网络的输出权值，能以极快的学习速度获得良好的推广性。为验证所提方法的有效性，将 KELM、KPLS、ALD-KRLS 用于不同实测交通流数据中，在同等条件下，与现有方法进行比较。实验结果表明，不同核学习方法的预测精度和训练速度均有提高，体现了核学习方法在短时交通流量预测中的应用潜力。

关键词：核学习方法；短时交通流；预测

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.09.0916

Short-term traffic flow forecasting based on kernel learning methods

Wang Qiuli, Li Jun

(*School of Automation & Electrical Engineering*, *Lanzhou Jiaotong University*, *Lanzhou* 730070, *China*)

Abstract: Based on the powerful nonlinear mapping ability of kernel learning, proposing a class of kernel learning method for the short-term traffic flow forecasting. Kernel recursive least squares (KRLS) method using Approximate Linear Dependence (ALD) technique can reduce the computational complexity and storage capacity, the KRLS method is an online kernel learning method and is suitable for training on large-scale data sets. Kernel partial least square (KPLS) method utilizes the covariance between input and output variables to extract latent features. Kernel extreme learning machine(KELM) method uses the kernel function to substitute for the unknown nonlinear feature mapping of the hidden layer, in addition, the output weights of the networks can also be analytically determined by using regularization least square algorithm, hence KELM method provides better generalization performance at a much faster learning speed. In order to verify the validity of the proposed kernel learning method, the employed ALD-KRLS, KPLS and KELM methods were respectively applied to different traffic flow forecasting instances in different area, compared to the other methods under the same conditions. Experimental results show that the proposed kernel-based learning methods have higher forecasting accuracy and have improved training speed in the short-term traffic flow forecasting.

Key Words: kernel learning methods; short-term traffic flow; forecasting

0　引言

交通流量预测作为智能交通系统(intelligent transportation systems, ITS)中的重要组成部分，在城市路网交通信号控制、拥塞控制、驾驶路线规划等方面发挥重要作用[1-2]。交通流量预测可为旅行者提供实时交通信息以节约出行时间与成本，更有助于交通决策部门制定积极主动的交通管理策略，以提高城市交通路网的运行效率。

神经网络与支持向量机(support vector machines，SVM)等机器学习方法近年来在短时交通流预测方面已取得许多成功应用。文献[3~5]将神经网络应用于交通流预测模型。文献[6~8]分别将 SVM、最小二乘支持向量机(LSSVM)应用于短时交通流单步预测中。极限学习机(ELM) [9]随机选择 SLFNs 的隐含层节点及相应参数，具有极快的学习速度，当 ELM 的隐含层节点未知时，采用核矩阵方法替代之，形成 KELM[11]方法，文献[10~13]成功将 ELM、KELM 应用于时间序列预测中。

考虑特征提取技术，文献[14，15] 分别将 SVM 与主元分析、核主元分析结合的 PCA-SVM、KPCA-SVM 方法应用于短

基金项目：国家自然科学基金资助项目（51467008）

作者简介：王秋莉（1992-），女，甘肃天水人，硕士研究生，主要研究方向为智能计算在时间序列预测中的应用（1205843513@qq.com）；李军（1969-），男，教授，硕导，博士，主要研究方向为计算智能与复杂非线性系统建模、控制.

时交通流预测中，在一定程度上均有效提高了预测精度。

文献[16]给出了 KPLS 方法，它是线性 PLS 方法在非线性特征空间上的延伸。考虑在线形式的核学习方法，文献[17]给出了 KRLS 方法，它基于稀疏技术限制核矩阵的规模，以适应较大数据集的训练。文献[18]将 KRLS 方法用于时间序列预测中，具有较好的预测效果。

鉴于 SVM、ELM 等方法在短时交通流预测等方面的成功应用，考虑到 KELM 等核学习方法在短时交通流预测领域的研究成果相对少见。针对上述问题，本文提出基于 KELM、KPLS、KRLS 的一类核学习预测方法，以进一步提高预测精度及提高计算效率。

为验证所提基于 KRLS、KELM、KPLS 方法的预测模型的有效性，将它们应用于某地区的实测数据实例中，在同等条件下，不仅对所给出的三种不同核学习方法进行比较，而且还将与现有 SVM 等预测方法进行实验对比，以进一步拓宽核学习方法在智能交通系统中的应用潜力。

1　核学习方法

给定训练数据 $\{x_i, y_i\} \in R^m \times R^n$，$i = 1,2,\ldots,N$，构建矩阵 $X \in \mathrm{R}^{N\times m}, Y \in \mathrm{R}^{N\times n}$。将数据 x_i 映射至高维特征空间中，即 $\phi: x \in R^m \to \phi(x) \in F \subseteq R^M$，为避免特征空间上高维计算的困难，考虑定义满足 Mercer 条件的核函数 $k(x_i, x_j) = \phi(x_i)^{\mathrm{T}} \phi(x_j)$，则有以 $k(x_i, x_j)$ 为元素构成的核矩阵 $K = \Phi^{\mathrm{T}}\Phi$，其中 $\Phi = [\phi(x_1)\ \phi(x_2) \ldots \phi(x_N)]$。

1.1　KELM 方法

KELM 是 ELM 方法在非线性特征空间上的延伸，首先考虑对于含 $\tilde{N}$ 个隐含层节点的 SLFNs，其网络输出为

$$y(\mathbf{x}) = \sum_{i=1}^{\tilde{N}} \beta_i h_i(\mathbf{x}) = \sum_{i=1}^{\tilde{N}} \beta_i G(\mathbf{x}; \mathbf{w}_i, b_i), \mathbf{x} \in \mathrm{R}^m, \boldsymbol{\beta}_i \in \mathrm{R}^n \quad (1)$$

其中：β_i 表示连接第 i 个隐含层节点与输出层节点的权值向量。$h_i(\mathbf{x})$ 表示第 i 个隐含层节点的输出函数 $G(\mathbf{x}; \mathbf{w}_i, b_i)$，由隐含层节点的类型确定。若隐含层节点为 RBF 节点，其激活函数形式为

$$G(\mathbf{x}; \mathbf{w}_i, b_i) = \exp\left(-b_i \left\|\mathbf{x} - \mathbf{w}_i\right\|^2\right) \quad (2)$$

若 SLFNs 能以零误差逼近 N 个训练样本，即 $\sum_{j=1}^{N} \left\|o_j - t_j\right\| = 0$，则存在 β_i，w_i 和 b_i 有如下表达式：

$$\sum_{i=1}^{\tilde{N}} \beta_i G(x_j; w_i, b_i) = t_j, j = 1,\ldots,N \quad (3)$$

矩阵形式可表示为

$$H\beta = Y \quad (4)$$

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(w_1 \cdot x_1 + b_1) & \cdots & G(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ G(w_1 \cdot x_N + b_1) & \cdots & G(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N\times\tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_{\tilde{N}}^{\mathrm{T}} \end{bmatrix}_{\tilde{N}\times n} \quad Y = \begin{bmatrix} y_1^{\mathrm{T}} \\ \vdots \\ y_N^{\mathrm{T}} \end{bmatrix}_{N\times n} \quad (5)$$

其中：H 为 SLFNs 的隐含层输出矩阵，H 的第 i 列表示与输入 $x_1, x_2, \ldots, x_N$ 相关的第 i 个隐含层节点的输出向量；第 i 行表示与输入 x_i 相关的隐含层特征映射。

若激活函数 $G(x_j; w_i, b_i)$ 在任意区间上无限可微，则当隐含层节点数目 $\tilde{N}$ 小于样本 N 时，SLFNs 仍能以极小的训练误差逼近训练样本[9]。此时矩阵 H 并非方阵，因此存在 $\hat{\beta}_i$，$\hat{w}_i$ 和 $\hat{b}_i$ 使得

$$\begin{aligned} &\left\|H(\hat{w}_1,\ldots,\hat{w}_{\tilde{N}}, \hat{b}_1,\ldots,\hat{b}_{\tilde{N}})\hat{\beta} - Y\right\| \\ &= \min_{w_i,\beta_i,b_i} \left\|H(w_1,\ldots,w_{\tilde{N}}, b_1,\ldots,b_{\tilde{N}})\hat{\beta} - Y\right\| \end{aligned} \quad (6)$$

因此，与一般的前馈神经网络方法调整各层网络权值的训练学习不同，ELM 方法仅需训练网络的输出权值矩阵 β，其学习过程相当于求解式(4)的方程组 $H\beta = Y$ 的最小二乘解 $\hat{\beta}$，即有

$$\hat{\beta} = H^{\dagger} Y \quad (7)$$

为进一步提高网络的泛化能力及矩阵求解的稳定性，根据岭回归思想引入正则化参数 ，对 β 的解可进行一定约束。因此，式(7)的求解可进一步表示为：

$$\hat{\beta} = H^T \left(HH^T + \frac{I}{\Upsilon}\right)^{-1} Y \quad (8)$$

相应地，ELM 网络的输出为

$$y(x) = h(x)\hat{\beta} \quad (9)$$

在式(8)(9)的求解过程中，进一步考虑用核函数表示未知的隐含层非线性特征映射，引入满足 Mercer 条件的核函数，定义核矩阵 $K = HH^{\mathrm{T}}$，其元素为 $[K]_{i,j} = k(x_i, x_j) = h(x_i) \bullet h(x_j)$。此时，KELM 网络输出为

$$\begin{aligned} y(x) &= h(x)\hat{\beta} = h(x)H^T\left(HH^T + \frac{I}{\Upsilon}\right)^{-1} Y \\ &= \begin{bmatrix} k(x, x_1) \\ \vdots \\ k(x, x_N) \end{bmatrix} \left(K + \frac{I}{\Upsilon}\right)^{-1} Y \end{aligned} \quad (10)$$

1.2　KPLS 方法

KPLS 将高维特征空间看做原始空间的对偶，基于核变换技术在对偶空间中建立线性 PLS 回归模型。PLS 方法利用输入 X 和输出 Y 矩阵的协方差选择潜在特征变量，其回归模型可写成矩阵形式：

$$Y = XB + F \quad (11)$$

其中：B 是维数为 $m \times n$ 的回归系数矩阵，F 为维数为 $N \times n$ 的残差矩阵。

文献[19]将线性 PLS 方法延伸至高维非线性特征空间，形成 KPLS 方法，其算法实现如下：

a)随机初始化输出矩阵 Y 的潜在向量 $\mathbf{u}$，并定义核矩阵 $K = \Phi^{\mathrm{T}}\Phi$。

b)计算输入矩阵 X 的潜在向量 t，并进行归一化处理，即：

$$t=\Phi^{\mathrm{T}}\Phi u=Ku,\qquad t=\frac{t}{\|t\|} \tag{12}$$

c)计算载荷向量 c。

$$c=Y^{\mathrm{T}}t \tag{13}$$

d)计算 u 并进行归一化处理，即：

$$u=Yc,\qquad u=\frac{u}{\|u\|} \tag{14}$$

e)重复 b)~e)，直至收敛。

f)对核矩阵 K 及输出矩阵 Y 进行收缩映射，即

$$K=\left(I-\mathbf{tt}^{\mathrm{T}}\right)\mathbf{K}\left(I-\mathbf{tt}^{\mathrm{T}}\right)=\mathbf{K}-\mathbf{tt}^{\mathrm{T}}\mathbf{K}-\mathbf{Ktt}^{\mathrm{T}}+\mathbf{tt}^{\mathrm{T}}\mathbf{Ktt}^{\mathrm{T}} \tag{15}$$

$$\mathbf{Y}=\mathbf{Y}-\mathbf{tt}^{\mathrm{T}}\mathbf{Y} \tag{16}$$

回归系数矩阵 $B=\Phi U(T^{\mathrm{T}}KU)^{-1}T^{\mathrm{T}}Y$。因此，多维输入、多维输出的 KPLS 回归模型在训练数据集上的预测输出为

$$\hat{\mathbf{Y}}=\Phi B=\mathbf{K}U(T^{\mathrm{T}}\mathbf{K}U)^{-1}T^{\mathrm{T}}Y=TT^{\mathrm{T}}Y \tag{17}$$

其中:矩阵 U 的各列由向量 u_j 构成，矩阵 T 的各列由向量 t_j 构成。

为消除回归模型中偏置项的影响，还应对映射到特征空间中的数据进行中心化处理，需要对训练数据及测试数据所构成的核矩阵分别进行如下变换，即：

$$\tilde{\mathbf{K}}=\left(\mathbf{I}-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}\right)\mathbf{K}\left(\mathbf{I}-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}\right) \tag{18}$$

$$\tilde{\mathbf{K}}_t=\left(\mathbf{K}_t-\frac{1}{N}\mathbf{1}_{N_t}\mathbf{1}_N^{\mathrm{T}}\mathbf{K}_t\right)\left(\mathbf{I}-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}\right) \tag{19}$$

其中: K_t 为未进行中心化的测试数据核矩阵，维数为 $N_t\times N$，N_t 为测试数据集的数据大小，$\mathbf{1}_N$ 是元素为 1，长度为 N 的列向量，$\mathbf{1}_{N_t}$ 是元素为 1，长度为 N_t 的列向量，I 为单位阵。

考虑用于时间序列预测建模的单输出情形，此时回归系数向量 $d=U(T^{\mathrm{T}}KU)^{-1}T^{\mathrm{T}}Y$，其元素为 $d_i,i=1,\ldots,N$。则 KPLS 预测模型的式(17)可进一步表示为

$$y(x)=\sum_{i=1}^{N}d_iK(x,x_i) \tag{20}$$

由式(17)，若仅提取前 p 个潜在变量，在非线性特征空间中，将 KPLS 预测模型解释为线性 PLS 回归模型，即

$$y(x)=c_1t_1(x)+\cdots+c_pt_p(x)=c^{\mathrm{T}}\mathbf{t}(x) \tag{21}$$

其中: $\{t_i(x)\}_{i=1}^{p}$ 是 x 的前 p 个潜在变量，即非线性主元上的投影。

1.3 KRLS 方法

在训练阶段，假设获取 t 时刻之前的训练数据 $(x_1,y_1),\ldots,(x_{t-1},y_{t-1})$ 后，由训练数据的子集可构成字典 $D_{t-1}=\{\tilde{x}_j\}_{j=1}^{s_{t-1}}$，其中 $\tilde{x}_j$ 是所选择的训练数据，s_{t-1} 是字典集合的数目。特征空间中，$\{\phi(\tilde{x}_j)\}_{i=1}^{s_{t-1}}$ 是线性独立的数据向量。当新数据 x_t 出现时，判断 $\phi(x_t)$ 是否满足 ALD 条件，若 $\delta_t>v$，则将 x_t 加入字典 D_t 中。

为考虑是否将数据 x_t 加入字典集合中，需求出满足 ALD 条件的系数向量 $a=(a_1,\ldots,a_{s_t-1})^{\mathrm{T}}$，即有

$$\delta_t=\min_{a}\left\|\sum_{j=1}^{s_{t-1}}a_j\phi(\tilde{x}_j)-\phi(x_t)\right\|^2\le v \tag{22}$$

其中阈值参数 $k(x_i,x_j)=\phi(x_i)^{\mathrm{T}}\phi(x_j)$ 的取值大小影响稀疏化程度。考虑 $k(x_i,x_j)=\phi(x_i)^{\mathrm{T}}\phi(x_j)$，由式(22)得到

$$\begin{aligned}\delta_t&=\min_{a}\Big\{\sum_{i,j=1}^{s_{t-1}}a_ia_jk(\tilde{x}_i,\tilde{x}_j)-\\&\quad 2\sum_{j=1}^{s_{t-1}}a_jk(\tilde{x}_j,x_t)+k(x_t,x_t)\Big\}\\&=\min\left\{a^{\mathrm{T}}\tilde{K}_{t-1}a-2a^{\mathrm{T}}\tilde{k}_{t-1}(x_t)+k_{tt}\right\}\end{aligned} \tag{23}$$

其中，$[\tilde{K}_{t-1}]_{i,j}=k(\tilde{x}_i,\tilde{x}_j)$，$(\tilde{k}_{t-1}(x_t))_i=k(\tilde{x}_i,x_t)$，$k_{tt}=k(x_t,x_t)$，$i,j=1,\ldots,s_{t-1}$。

求解式(23)，可得到最优系数向量 a 及 ALD 条件：

$$a_t=\tilde{K}_{t-1}^{-1}\tilde{k}_{t-1}(x_t),\qquad \delta_t=k_{tt}-\tilde{k}_{t-1}(x_t)^{\mathrm{T}}a_t\le v \tag{24}$$

借助 ALD 条件，直至 t 的所有数据均能由字典集合 D_t 中数据向量的近似线性组合来表示。由式(22)可推得

$$\phi(x_i)=\sum_{j=1}^{s_i}a_{i,j}\phi(\tilde{x}_j)+\phi_i^{\mathrm{res}},\ \|\phi_i^{\mathrm{res}}\|^2\le v \tag{25}$$

其中 ϕ_i^{res} 为残差成分向量，$a_{i,j}$ 是系数向量 a_i 的第 j 个元素。延伸至特征空间，定义矩阵 $\tilde{\Phi}=[\phi(\tilde{x}_1)\ \phi(\tilde{x}_2)\ldots\phi(\tilde{x}_m)]$，$\Phi^{\mathrm{res}}=[\phi(x_1^{\mathrm{res}})\ \phi(x_2^{\mathrm{res}})\ldots\phi(x_N^{\mathrm{res}})]$。式(25)的矩阵表达为

$$\Phi=\tilde{\Phi}\mathrm{A}^{\mathrm{T}}+\Phi^{\mathrm{res}} \tag{26}$$

当 v 足够小时，式(26)的近似表达式为 $K_t\approx \mathrm{A}_T\tilde{K}_t\mathrm{A}_t^{\mathrm{T}}$，其中 $[K_t]_{i,j}=k(x_i,x_j)$，$[\mathrm{A}_t]_{i,j}=a_{i,j}$，$i,j=1,\ldots,t$。

在时刻 $L(w)=\sum_{i=1}^{t}\left(f(x_i)-y_i\right)=\|\Phi_t^{T}w-Y_t\|^2$，由 RLS 算法可知，最小化误差平方和满足下式：

$$L(w)=\sum_{i=1}^{t}\left(f(x_i)-y_i\right)=\|\Phi_t^{T}w-Y_t\|^2 \tag{27}$$

其中矩阵 $Y_t=(y_1;\ldots;y_t)$，维数为 $t\times n$。w 为权值系数向量。$\|\cdot\|^2$ 表示 2 范数。由式(27)得 w 的最优解为

$$w_t=\sum_{i=1}^{t}\alpha_i\phi(x_i)=\Phi_t\alpha \tag{28}$$

其中 $\alpha=(\alpha_1,\ldots,\alpha_t)^{\mathrm{T}}$，由式(26)(27)得到

$$L(\alpha)=\|K_t\alpha-Y_t\|^2 \tag{29}$$

式(29)进行稀疏化处理，令字典数据向量形成的核矩阵 $\tilde{K}_t$ 逼近 K_t，$\tilde{K}_t$ 是由字典集合中的所有 s_t 个元素构成的核矩阵，借助 ALD 条件从训练数据中构建字典集合。

式(26)代入式(28)，则 $w_t = \Phi_t \alpha_t \approx \tilde{\Phi}_t A_t^T \alpha_t = \tilde{\Phi}_t \tilde{\alpha}_t$，其中 $\tilde{\alpha}_t = A_t^T \alpha_t$ 为 st 个“约简”系数构成的向量。因此，式(29)变形为

$$L(\tilde{\alpha}) = \left\| \Phi_t^T \tilde{\Phi}_t \tilde{\alpha} - Y_t \right\|^2 = \left\| A_t \tilde{K}_t \tilde{\alpha} - Y_t \right\|^2 \tag{30}$$

式(30)的最小化求解为

$$\tilde{\alpha}_t = \left(A_t \tilde{K}_t\right)^{\dagger} Y_t = \tilde{K}_t^{-1}\left(A_t^T A_t\right)^{-1} A_t^T Y_t \tag{31}$$

多维输入-输出 ALD-KRLS 方法的具体实现如下：

a)训练阶段，给定阈值参数 v。时刻 t=1 时，获取数据 (x_1, y_1)，$\tilde{K}_1 = [k_{11}]$，$\tilde{K}_1^{-1} = [1/k_{11}]$，$\alpha_1 = (y_1/k_{11})$，$P_1 = [1]$，$s = 1$。

b)令 t=t+1，得到新数据 (x_t, y_t)，计算

$$\tilde{k}_{t-1}(x_t) = \left[k(\tilde{x}_1, x_t), k(\tilde{x}_2, x_t), \ldots, k(\tilde{x}_{s_{t-1}}, x_t)\right]^T \tag{32}$$

c)计算式(24)，如果 $\delta_t > v$，将 x_t 加入字典中，字典矩阵 $\tilde{K}_t$ 也相应增长。即有 $D_t = D_{t-1} \cup \{\tilde{x}_t\}$，$s = s+1$，更新计算 $\tilde{K}_t^{-1}$：

$$\begin{gathered} \tilde{K}_t = \begin{bmatrix} \tilde{K}_{t-1} & \tilde{k}_{t-1}(x_t) \\ \tilde{k}_{t-1}(x_t)^T & k_{tt} \end{bmatrix} \Rightarrow \\ \tilde{K}_t^{-1} = \frac{1}{\delta_t}\begin{bmatrix} \delta_t \tilde{K}_{t-1}^{-1} + a_t a_t^T & -a_t \\ -a_t^T & 1 \end{bmatrix} \end{gathered} \tag{33}$$

由于 x_t 是字典中的数据向量，相应有：

$$A_t = \begin{bmatrix} A_{t-1} & 0 \\ 0^T & 1 \end{bmatrix}, A_t^T A_t = \begin{bmatrix} A_{t-1}^T A_{t-1} & 0 \\ 0^T & 1 \end{bmatrix} \tag{34}$$

P_t 更新为

$$P_t = \left(A_t^T A_t\right)^{-1} = \begin{bmatrix} P_{t-1} & 0 \\ 0^T & 1 \end{bmatrix} \tag{35}$$

其中 $\mathbf{0}$ 是零向量。

由式(24)得到 $\alpha_t^T \tilde{K}_{t-1} = \tilde{k}_{t-1}(x_t)^T$，因此 $\tilde{\alpha}$ 更新为

$$\begin{aligned} \tilde{\alpha}_t &= \tilde{K}_t^{-1}\left(A_t^T A_t\right)^{-1} A_t^T Y_t \\ &= \tilde{K}_t^{-1}\begin{bmatrix} \left(A_{t-1}^T A_{t-1}\right)^{-1} A_{t-1}^T Y_{t-1} \\ q_t \end{bmatrix} \\ &= \begin{bmatrix} \tilde{\alpha}_{t-1} - \dfrac{a_t}{\delta_t}\left(q_t - \tilde{k}_{t-1}(x_t)^T \tilde{\alpha}_{t-1}\right) \\ \dfrac{1}{\delta_t}\left(q_t - \tilde{k}_{t-1}(x_t)^T \tilde{\alpha}_{t-1}\right) \end{bmatrix} \end{aligned} \tag{36}$$

输出字典 D_t，$\tilde{\alpha}_t$。

d)如果 $\delta_t \le \mu$，$D_t = D_{t-1}$，$s = s$，$\tilde{K}_t = \tilde{K}_{t-1}$。更新 $A_t = \left[A_{t-1}^T, a_t\right]^T$。按矩阵求逆引理更新 P_t：

$$P_t = P_{t-1} - \frac{P_{t-1} a_t a_t^T P_{t-1}}{1 + a_t^T P_{t-1} a_t} \tag{37}$$

若定义 $\beta_t = \dfrac{P_{t-1} a_t}{1 + a_t^T P_{t-1} a_t}$，则 $\beta_t = P_t a_t$，$\tilde{k}_{t-1}(x_t) = \tilde{K}_t a_t$。更新计算 $\tilde{\alpha}$ 如下：

$$\begin{aligned} \tilde{\alpha}_t &= \tilde{K}_t^{-1} P_t A_t^T Y_t \\ &= \tilde{K}_t^{-1}\left(P_{t-1} - \beta_t a_t^T P_{t-1}\right)\left(A_{t-1}^T Y_{t-1} + a_t q_t\right) \\ &= \tilde{\alpha}_{t-1} + \tilde{K}_t^{-1}\left(P_t a_t q_t - \beta_t a_t^T \tilde{K}_t \tilde{\alpha}_{t-1}\right) \\ &= \tilde{\alpha}_{t-1} + \tilde{K}_t^{-1} \beta_t \left(q_t - \tilde{k}_{t-1}(x_t)^T \tilde{\alpha}_{t-1}\right) \end{aligned} \tag{38}$$

输出字典 D_t，$\tilde{\alpha}_t$。

e)迭代计算 Step 2~ Step 4，直至所有训练数据依次完成。

f)测试阶段，对已训练的 KRLS 模型进行测试，其预测输出如式(39)所示。

$$\hat{y}_t = \tilde{k}_{t-1}(x_t)^T \tilde{\alpha}_{t-1} \tag{39}$$

若训练数据的数目为 N，则 KRLS 算法的计算复杂度为 $O(Nm^2)$。

2　短期交通流预测实例

按照时间序列建模的方式，将本文的不同核学习方法应用于短期交通流预测实例中，所建立的预测模型为

$$\hat{y}(t+h) = f(x_t) \tag{40}$$

其中:h 表示预测步长，x_t 是历史时间序列值构成的多维输入向量 $x_t = [y_{t-m}, \cdots y_{t-1}]$，$m$ 为嵌入维数。对于单变量交通流时间序列，构建训练数据对 $\{x_i, y_i\} \in R^m \times R$，$i = 1, 2, \ldots, N$，此时模型输出维数 n=1，$f(\cdot)$ 则分别对应于 KRLS、KPLS、KELM 方法，并在同等条件下，与 KPCA-SVM、SVM 等方法对比。

实验采用正则化均方根误差(normalized mean square root error, NRMSE)以及均方根误差(mean square root error, RMSE)、平均绝对百分比误差(mean absolute percentage error，MAPE) 作为衡量指标。NRMSE 表达式如下：

$$\mathrm{NRMSE} = \left\{ \frac{\sum_{i=1}^{N_t} (y - \hat{y})^2}{(N_t - 1)\operatorname{var}(y)} \right\}^{1/2} \tag{41}$$

其中: $\hat{y}$ 表示模型的预测输出，y 为实际输出，$\operatorname{var}(y)$ 表示序列方差，N_t 为测试数据数目。实验中的高斯核函数形式如下:

$$k(x_i, x_j) = \exp\left\{ -\left\| x_i - x_j \right\| / 2\sigma^2 \right\} \tag{42}$$

其中: σ 为高斯核函数的宽度。

2.1　实例 1

选取北京某公路观测站的实测交通流数据，观测总时段为 4 天共 96 小时，以 15 分钟为采样间隔，记录了 384 组数据。实验中取前 285 组数据为训练数据，后 92 组数据为测试数据，选取嵌入维数 m=4。由交叉验证法确定不同核学习方法的预测模型参数。KELM ，KPLS， KRLS，KPCA 均各自选取高斯核函数，核函数宽度 σ=1。KELM 正则化参数 γ=50；KPLS 中潜在变量数目 p=15；ALD-KRLS 最大字典容量 s_{max}=200，阈值 v=0.1。对比实验中，SVM 方法使用 Libsvm 软件完成，置惩罚因子 $C = 0.5$，$\varepsilon = 0.01$；KPCA-SVM 方法中，选取非线

性主元数目为 15，SVM 的核函数为线性核函数。结果评价由表 1、2 列出。

表 1　不同核学习方法的提前 15 min 的预测结果对比

预测方法	MAPE	RMSE	NRMSE
SVM	33.808 6	23.454 8	0.269 3
KPCA-SVM	33.910 2	23.334 6	0.267 9
KELM	33.462 9	23.260 7	0.267 7
KPLS	33.443 7	22.582 2	0.259 3
KRLS	33.860 6	22.521 0	0.250 6

表 2　不同核学习方法的提前 30 min 的预测结果对比

预测方法	MAPE	RMSE	NRMSE
SVM	43.446 6	30.429 8	0.347 6
KPCA-SVM	43.981 9	30.021 5	0.343 0
KELM	45.288 8	29.867 9	0.341 2
KPLS	47.131 0	29.282 5	0.334 5
KRLS	42.024 6	28.254 5	0.314 4

由表 1、2 可看出， KELM、KPLS、KRLS 在交通流量预测中均取得了较好的预测效果，且 KRLS 预测效果最佳。

图 1、2 分别给出了不同核学习方法在测试集上进行提前 15 min、30 min 预测的对比曲线。由图 1、2 看出，不同核学习方法均能准确地预测实际交通流量值，预测误差波动相对较小，呈现出好的预测效果。图 3 还给出 KRLS 方法进行单步预测时，在训练数据集上其均方误差 MSE 的变化曲线，由图可知，KRLS 方法具有较快的收敛速度。

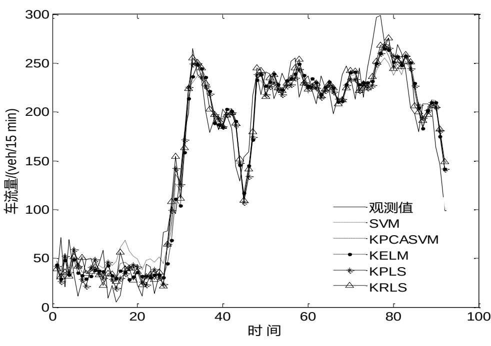


图 1　基于核学习方法的提前 15 min 的预测结果

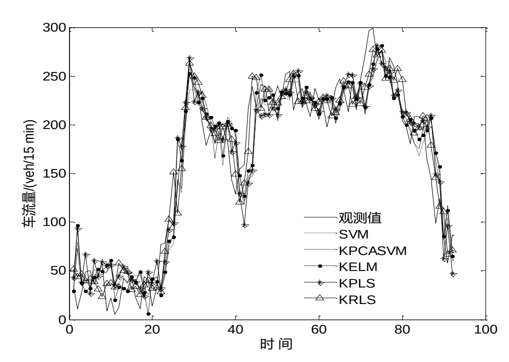


图 2　基于核学习方法的提前 30 min 的预测结果

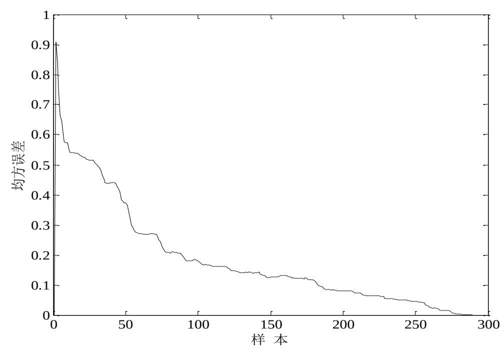


图 3　训练集上 KRLS 方法的收敛曲线

2.2　实例 2

实例选取英国某地区交通局提供的交通流量数据，采样间隔为 15min，采样时间为 2011 年 3 月，本实验截选了 336 个时间点。选取嵌入维数 m=3。实验中取前 235 个数据作为训练集输入，最后 100 个数据作为测试样本。

参数设定如下所示。KELM ，KPLS， KRLS，KPCA，SVM 均各自选取高斯核函数，核函数宽度 $\sigma=1$。KELM 正则化参数 $\gamma=100$；KPLS 中潜在变量数目 $p=20$；ALD-KRLS 最大字典容量 $s_{\max}=90$，阈值 $\nu=0.1$。对比实验中，SVM 方法惩罚因子 $C=0.5$, $\varepsilon=0.01$；KPCA-SVM 方法中，KPCA 的非线性主元数目为 10，SVM 选取线性核函数。

表 3　不同核学习方法的提前 15 min 的预测结果对比

预测方法	MAPE	RMSE	NRMSE
SVM	16.454 3	66.558 7	0.698 7
KPCA-SVM	15.912 0	62.629 9	0.626 1
KELM	15.698 1	61.408 0	0.645 2
KPLS	15.389 7	58.458 3	0.587 8
KRLS	13.885 6	55.012 2	0.492 0

表 4　不同核学习方法的提前 30 min 的预测结果对比

预测方法	MAPE	RMSE	NRMSE
SVM	20.520 9	76.084 4	0.770 4
KPCA-SVM	19.665 1	74.977 8	0.683 9
KELM	18.989 5	72.168 7	0.664 8
KPLS	17.163 0	69.615 8	0.621 0
KRLS	15.652 1	65.478 9	0.548 2

图4、5 分别为不同核学习方法分别进行提前 15 min、30 min 的单步、多步预测时在测试集上的预测结果对比曲线，由图 4、5 的结果看出，本文的核学习方法呈现出较好的预测效果。

在测试集上，不同核学习方法分别进行提前 15 min、30 min 预测的预测结果评价指标对比则由表 3、表 4 给出。从表 3、表 4 可以看出，与 SVM 方法相比，KELM 方法的预测精度略有提升，KPLS 方法次之，KRLS 方法则显示出了较高的预测精度。通过算法训练的收敛时间对比可知，KELM 与 KRLS 方法的算法收敛速度较快；KPLS 预测方法的训练速度较慢；ALD-KRLS

方法具有自适应在线训练特性，其收敛速度快，预测精度最高。

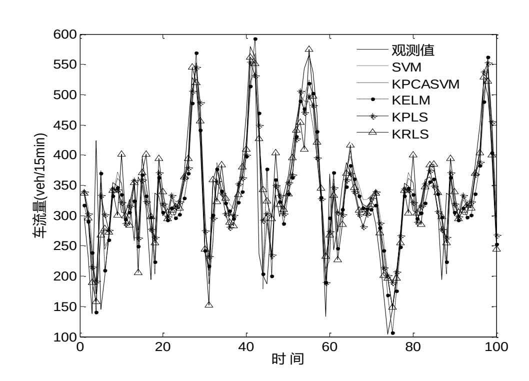


图 4 基于核学习方法的提前 15 min 的预测结果

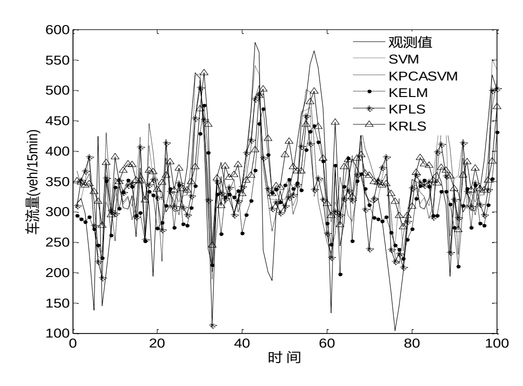


图 5 基于核学习方法的提前 30 min 的预测结果

3 结束语

针对交通流短期预测，分别提出基于 KELM、KPLS、KRLS 的核学习预测模型，KELM 方法以核映射代替了传统 ELM 的随机映射，提高了网络的泛化能力和学习速度。KPLS 方法具有特征提取能力强、算法实现容易等优点。KRLS 方法具有时变自适应学习特性，通过稀疏化算法，可控制核矩阵的大小，从而提高计算效率。实验结果表明，将不同核学习方法应用于短期交通流预测中是一种可行且有效地选择，另一方面，KELM 与 KPLS 方法均适宜于离线训练，其不足在于当训练数据集的数据超过一定规模时，会引起核矩阵的运算困难。KRLS 方法则适宜于在线训练，使其能用于较大规模数据集的训练。

参考文献：

[1] Lippi M, Bertini M, Frasconi P. Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning [J]. IEEE Trans on Intelligent Transportation Systems, 2013, 14 (2): 871-882.

[2] Sun S, Zhang C, Yu G. A bayesian network approach to traffic flow forecasting [J]. IEEE Trans on Intelligent Transportation Systems, 2006, 7 (1): 124-132.

[3] Messer C, Thomas U I. Short-term freeway traffic volume forecasting using radial basis function neural network [J]. Transportation Research Record Journal of the Transportation Research Board, 1998, 1651 (1): 39-47.

[4] 李军，黄杰．基于自组织映射神经网络的局部自回归方法在网络流量预测中的应用 [J]. 信息与控制, 2016, 45 (1): 120-128.

[5] 许榕，周东，蒋士正，等．自适应粒子群神经网络交通流预测模型 [J]. 西安交通大学学报, 2015, 49 (10): 103-108.

[6] Zhang Y L, Xie Y C. Forecasting of short term freeway volume with support vector machines [J]. Transportation Research Record, 2007, 12 (2): 92-99.

[7] 陈小波，刘祥，等．基于 GA-LSSVR 模型的路网短时交通流预测研究 [J]. 交通运输系统工程与信息, 2017, 17 (1): 60-73.

[8] 商强，杨兆升，等．基于奇异谱分析和 CKF-LSSVM 的短时交通流量预测 [J]. 吉林大学学报：工学版, 2016, 46 (6): 1792-1798.

[9] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feed forward neural networks [C]// Proc of IEEE International Joint Conference on Neural Networks. 2004, 2: 985-990.

[10] 商 强，杨兆升，李志林，等．基于相空间重构和 RELM 的短时交通流量预测 [J]. 华南理工大学学报：自然科学版, 2016, 44 (4): 109-114.

[11] Huang G B, Wang D H, Lan Y. Extreme learning machines: a survey [J]. International Journal of Machine Learning and Cybernetics, 2011, 2 (2): 107-122.

[12] 李军，李大超．基于优化核极限学习机的风电功率时间序列预测 [J]. 物理学报, 2016, 65 (13): 33-42.

[13] 马超，张英堂．在线核极限学习机及其在时间序列预测中的应用 [J]. 信息与控制, 2014, 43 (5): 624-629.

[14] Chen Q, Chen X, Wu Y. Optimization algorithm with kernel PCA to support vector machines for time series prediction [J]. Journal of Computers, 2010, 5 (3) .

[15] Sun Z, Fox G. Traffic flow forecasting based on combination of multidimensional scaling and SVM [J]. International Journal of Intelligent Transportation Systems Research, 2014, 12 (1): 20-25.

[16] Rosipal R, Trejo L. Kernel partial least squares regression in reproducing kernel Hilbert space [J]. Journal of Machine Learning Research, 2001, 2 (2): 97-123.

[17] Engel Y, Mannor S, Meir R. The kernel recursive least-squares algorithm [J]. IEEE Trans on Signal Processing, 2004, 52 (8): 2275-2285.

[18] Fan H, Song Q. A sparse kernel algorithm for online time series data prediction [J]. Expert systems with applications, 2013, 40 (6): 2174-2181.

[19] 李军，董海鹰．基于小波核偏最小二乘回归方法的混沌系统建模研究 [J]. 物理学报, 2008, 57 (8): 4756-4765.